ARL-TR-8383 ● June 2018

US Army Research Laboratory

Correlation between Identification Accuracy and Response Confidence for Common Environmental Sounds

by Kelly Dickerson, Ashley Foots, Alecia Moser, and Jeremy Gaston

ARL-TR-8383 ● JUN 2018

US Army Research Laboratory

Correlation between Identification Accuracy and Response Confidence for Common Environmental Sounds

by Kelly Dickerson, Ashley Foots, and Jeremy Gaston
Human Research and Engineering Directorate, ARL

Alecia Moser
Binghamton University, Department of Psychology

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
June 2018	Technical Report	1 May 2017–1 October 2018

4. TITLE AND SUBTITLE	
Correlation between Identification Accuracy and Response Confidence for Common Environmental Sounds	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Kelly Dickerson, Ashley Foots, Alecia Moser, and Jeremy Gaston	ARL 16–110
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
US Army Research Laboratory ATTN: RDRL-HRF-D Aberdeen Proving Ground, MD 21005-5425	ARL-TR-8383
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

One of the difficulties in studying environmental sound perception is determining what factors lead to successful identification. Environmental sounds as a class cover a broad range of acoustic and semantic attributes, thus challenging researchers who seek to balance the need for a representative set of environmental sounds with stimulus control and precision. The present study is one in a series of efforts to provide a baseline evaluation of a set of environmental sounds that is representative of the everyday environment. Fifteen listeners were presented with 41 different environmental sounds from six broad categories: household items, alarms, animals, human generated, mechanical, and vehicle sounds. Each sound was presented five times and participants had to generate a label for each of the samples. After typing their response, participants were then asked to rate their confidence in the accuracy of the label on a 7-point Likert scale. Participants were most accurate labeling alarms and human-generated sounds, consistent with previous studies using similar categories. Further, participants' confidence was highest when an accurate label was provided, suggesting that feelings of uncertainty were related to the ability to generate an accurate label.

15. SUBJECT TERMS

environmental sound perception, urban sounds, subjective ratings, Likert scales, confidence ratings, stimulus uncertainty

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Kelly Dickerson
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	21	410-278-5979

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction and Background

Auditory identification, the ability to associate meaning (i.e., retrieval of experience-based semantic knowledge) with a sound sample, assessed via the accurate selection or generation of a linguistic label for a signal, is a process that involves recognition of a particular pattern of auditory stimulation and linking that information to a sound-producing event stored in memory (Ballas and Howard 1987). The ability of a listener to link a signal to its source, thereby correctly identifying the signal, is determined by a combination of factors that fall broadly into two types of attributes: physical (i.e., acoustic) and content (i.e., semantic) features (Bergman et al. 2009; Gygi et al. 2007). While the literature agrees strongly that both acoustic and semantic information influence environmental sound identification (Lemaitre and Heller 2013; Lemaitre et al. 2013; Gregg and Samuel 2009), the weight of each of these factors, and how task context might shift the influence of either of these features, is still largely unknown.

There is some suggestion that the assumption auditory perception is based largely on acoustic characteristics is driven by research methodologies that bias participants toward using an acoustically driven strategy. For example, Truax (2001) argues that analytical listening strategies encourage listeners to focus on acoustic details while everyday listening is a task focused on increasing listeners' awareness of the contents of their environment; that is, everyday listening emphasizes the identity or meaning of a sound event rather than the basic acoustic properties of the event—a listener hears a car pass by, not a continuous lower-frequency broadband sound that increases in intensity and frequency as it changes position. This idea, that everyday listening is semantically oriented, is supported by sound-classification studies that suggest the potential for an environmental sound lexicon that mirrors the semantic organizational structure for meaningful speech. For example, like speech, identification of environmental sounds depends on frequency and familiarity but also a variety of other factors specific to environmental sounds, such as contextual congruency (Leech et al. 2009), the concreteness of the sound (Lemaitre et al. 2013), and the ease with which a sound can be described (Giordano et al. 2010). Ballas (1993) used listener surveys to gauge the frequency of occurrence for common environmental sounds; the ecological frequency surveys demonstrate that identification time and accuracy were directly related to the frequency with which a sound occurs in the environment. Van Petten and Rheinfelder (1995) and more recently Cummings et al., (2006) extended the results of Ballas (1993) by demonstrating similar accuracy responses and similar event-related potentials (ERPs) for speech and meaningful environment. Specifically, both studies found that spoken words and environmental sounds elicited the N400,[*] indicating processing at the semantic level, differing only in the overall amplitude of the waveforms. These findings suggest speech and sound may have slightly different neural-source generators but shared spatial regions of processing. An in-depth discussion of the N400 is beyond the scope of this paper but, in general, these results support the use of identification performance for environmental sounds as a metric of the ease with which a linguistic descriptor can be applied to the sound, supporting the argument that environmental sounds are perceived as semantic objects.

2. Present Study

The current study uses an open-ended response method to evaluate identification accuracy and confidence ratings for a set of 41 common environmental sounds. The 41 sounds selected for inclusion in this study fall into one of six experimenter-defined categories: household related, alarms, animals, human generated (nonspeech), mechanical, and vehicle sounds. These categories were selected based on their consistency with other research aimed at defining categories of environmental sounds (Gygi et al. 2007; Houix et al. 2012; Marcell et al. 2000) and are consistent with a representative urban environment based on Ballas (1993) and our own assessment of ecological frequency (Foots et al. 2016; McArdle et al. 2017). The open-ended identification task was selected instead of the classical multiple-alternative forced choice (XAFC) task, because there is some evidence this method may be more appropriate for measuring environmental sound identification as it provides more information about what cues a listener is using to make their identification judgment. Previous research from VanDerveer (1979), Ballas (1993), and Van Petten and Rheinfelder (1995) using the open-ended identification method for evaluation of semantic objects established that this type of metric can be used over the XAFC method as a strategy for gaining insight into not only the accuracy of the perceptual decision, but also the strategy used by the listener to reach that decision. For example, reaction time, word length effects, and lexical analyses can be conducted on open-ended response data (Duff et al., forthcoming 2018). Additionally, allowing open-ended responses enables for a larger number of stimulus evaluations than the XAFC methodology.

The primary purpose of this study is to evaluate the identifiability of a set of environmental sounds to be used in the creation of auditory arrays in subsequent experiments (Gygi and Shafiro [2009, 2010] discuss the growing recognition of the importance of norming). We expect to observe a range of accuracies, which will

[*] N400 has been defined as "[t]he component of the ERP which has been mostly closely tied to language processing … a late negative wave peaking at about 400 msec post-stimulus onset" (Van Patten and Rheinfelder 1995).

vary as a function of stimulus category. Specifically, as has been previously reported, we expect sounds generated from living sources (humans and animals) will be identified with greater accuracy than sounds generated from nonliving sources (vehicles and mechanical) with the likely exception of alarms, which for most adults are highly meaningful and tend to capture attention effectively. (Stavropoulos and Carver [2016] discuss living versus nonliving sounds, while Catchpole and McKeown [2017] provide overview of auditory alarms.)

3. Methods

3.1 Participants

Fifteen undergraduate students volunteered to serve as listeners in this study. All participants passed a hearing screening, defined as correctly responding to pure tone signals presented at intervals of 500, 1000, 2000, 4000, and 8000 Hz at 25 dB(HL) on a GSI Arrow 1800 audiometer with a TDH 39 headset. All participants provided informed consent and received course credit for their participation in the study.

3.2 Stimuli

Table 1 contains a list of the 41 environmental sounds used in the current study. Thirty-one of the sounds were downloaded from freesound.org, an open-access, user-supported sound library; four were recorded specifically for inclusion in this study; and six were samples selected from an existing sound library at the US Army Research Laboratory. Table 1 also lists an experimenter-defined category assigned to each of the sounds. All sounds were normalized for duration and level in Adobe Audition (CS 6 v. 5.0.2). Specifically, each sound was shortened to 1000 ms, including 5-ms linear on–off ramps. In addition, sounds were normalized for root-mean-square amplitude through a batch process to minimize potential loudness differences among sounds.

Table 1 Average accuracy, standard-error (SE), and confidence ratings for each of the 41 stimuli presented to listeners

Stimulus	Category	Accuracy	SE	Confidence	SE
Dishes1	Household	0.689	0.108	5.643	0.382
Metal1	Household	0.101	0.051	4.173	0.367
Pouring1	Household	1.000	0.000	6.707	0.136
ShakingCans1	Household	0.033	0.027	3.457	0.350
ShovelScrape1	Household	0.029	0.018	4.690	0.428
Waves1	Household	0.217	0.090	4.280	0.325
Alarm1	Alarm	0.827	0.089	6.120	0.244
Bell1	Alarm	1.000	0.000	6.627	0.223
Bell2	Alarm	0.987	0.013	6.703	0.148
Cellphone1	Alarm	0.920	0.067	6.613	0.158
Crickets1	Animal	0.813	0.079	6.267	0.238
Crickets2	Animal	0.947	0.041	6.680	0.140
Dog1	Animal	0.986	0.013	6.880	0.093
Dog2	Animal	0.293	0.112	5.973	0.195
Dog3	Animal	0.493	0.122	5.274	0.359
Baby1	Human	1.000	0.000	6.800	0.200
Baby2	Human	1.000	0.000	6.760	0.226
Walking1	Human	0.726	0.102	5.343	0.347
Walking2	Human	0.840	0.081	5.710	0.326
Bike1	Mechanical	0.507	0.128	4.673	0.494
Bike2	Mechanical	0.471	0.127	4.693	0.420
Jackhammer1	Mechanical	0.760	0.107	6.113	0.274
Lighter1	Mechanical	0.757	0.105	5.947	0.348
Shopvac1	Mechanical	0.824	0.089	6.547	0.183
Shopvac2	Mechanical	0.507	0.106	5.530	0.311
Bus1	Vehicle	0.528	0.104	5.357	0.323
Bus2	Vehicle	0.534	0.113	5.387	0.288
Bus3	Vehicle	0.397	0.100	4.497	0.350
Bus4	Vehicle	0.486	0.115	5.363	0.298
Helicopter1	Vehicle	0.680	0.108	5.740	0.307
Helicopter2	Vehicle	0.853	0.074	5.893	0.265
Motorcycle1	Vehicle	0.288	0.057	5.073	0.242
Motorcycle2	Vehicle	0.573	0.101	5.750	0.215
Plane1	Vehicle	0.613	0.118	5.857	0.241
Plane2	Vehicle	0.493	0.121	4.337	0.445
Plane3	Vehicle	0.413	0.118	5.547	0.314
Tank1	Vehicle	0.320	0.073	5.203	0.266
Truck1	Vehicle	0.773	0.070	5.807	0.323
Truck2	Vehicle	0.473	0.072	5.120	0.191
Truck3	Vehicle	0.722	0.086	5.547	0.228
Van1	Vehicle	0.149	0.044	4.483	0.341

3.3 Materials and Apparatus

All sounds were presented to listeners over Beyerdynamics T 70 closed reference headphones at a comfortable experimenter set listening level of 70dB(C). The presentation, timing, and recording of participant responses were controlled by E-Prime experiment development platform (Psychology Software 2012). The E-Prime experiment was run on a standard desktop computer.

3.4 Procedures

Following informed consent, listeners completed a hearing screening and then moved to the experimental computer and donned the headphone set. Each of the 41 sounds was presented to listeners five times (205 trials per participant) and the order of stimulus presentation was fully randomized. On each trial, the listener would press the spacebar to begin; the prompt "listen" would display 500 ms before stimulus onset and was displayed for the duration of the sound presentation (1000 ms). At the end of the sound, a text field appeared and participants typed their identification response using the computer keyboard. Participants were not given any specific instructions to shape the specificity of their response; they were simply instructed to use the label that made the most sense to them. Following the identification response, listeners were then prompted to rate their confidence in their identification label using a 1–7 Likert-type scale where "7" was highly confident and "1" was highly unconfident. Trials were self-paced, and a new trial began only after the participant initiated the trial by pressing the spacebar. Listeners were instructed to keep their responses brief and to restrict responses to a single word or short phrase; however, there was no limit on the number of characters that could occur in the participant-response field. The brevity instruction was meant to reduce the complexity of the response-classification task for subsequent coding by experimental raters.

3.5 Data Processing

When listeners are allowed to freely identify, the determination of the correctness of that response must be operationally defined to avoid the possibility of biased scoring. The current study adopted the same response-classification criteria as VanDerveer (1979), where a response was considered correct if it referred to the generating event or to a class of events that would include the generating event. For example, when responding to the sound "helicopter", responses such as "blades chopping the air" or "blades moving through air" would be considered correct (in addition to simply providing the label helicopter). Alternatively, responses such as "lawn mower" or "chopping noise" while describing similar events would be near misses and labeled as incorrect responses. Text responses were evaluated by two independent raters and classified as either a correct or incorrect response. Rated responses were then converted to either a one or zero so that *proportion correct* could be calculated for each sound event. The proportion correct was also calculated for items belonging to one of the six associated categories listed in Table 1 to examine an overall proportion correct for each sound category.

3.6 Inter-rater Reliability

Both raters evaluated all of the data produced by the entire sample of 15 participants. The Pearson correlation coefficient for the relationship between the two raters' scores was $r = 0.97$. However, Pearson correlation measures of inter-rater reliability for binary variables are often elevated by the high likelihood that the raters will agree simply based on chance. To control for this possibility, inter-rater reliability was also evaluated using Cohen's Kappa, which corrects for agreement by chance. The raters in this study produced a kappa score ($\kappa = 0.75$) that was significantly greater than chance (p <.001). This kappa score reflects substantial agreement between raters (Cohen 1960), and this level of agreement between raters means the criteria for recoding the open-ended responses as binary accuracy data were both clear and stable across the 3,075 trials included in the inter-rater reliability evaluation. (The reliability raters were authors Dickerson and Foots; on trials where the raters disagreed, Dickerson was the deciding authority with scores produced by Dickerson as the point of comparisons.)

4. Results and Discussion

All of the analyses that follow are based on the converted responses and confidence ratings. The identification responses were classified as either correct or incorrect and represent a binary measure of accuracy.

4.1 Identification Accuracy

Proportion correct was computed for each stimulus item for each participant. The total number of presentations of each stimulus (denominator in proportion-correct calculation) was adjusted to remove trials where no response was provided. This data-cleaning resulted in a loss of only 35 trials, distributed evenly across participants and stimulus items, reflecting a loss of only 1.8% of the total trials. This data-cleaning ensured the reported analyses only reflect responses made by participants and not the absence of response, which could have occurred for a variety of reasons.

Overall, average identification accuracy varied substantially across each of the 41 sounds in the sample (*M* = 0.61, *SE* = 0.04, *Min* = 0.02, *Max* = 1.00; see Table 1). For subsequent analyses we grouped the 41 stimuli into six meaning-based categories: household items (N = 6), alarms (N = 4), animal/insect (N = 5), human-generated nonspeech (N = 4), nonvehicle mechanical sounds (N = 6), and vehicles (N = 16). These six experimenter-defined categories and the number of items within each category were created based on a previous study (Foots et al. 2016; McArdle et al. 2017) where ecological frequency for environmental sounds was measured based on sampling from a real urban environment in the greater Baltimore, Maryland, area. The heavy representation of vehicles in this sound set is intentional and representative of most adults' everyday environment. The means for each of the six experimenter-defined categories are depicted in Fig. 1.

To examine the differences in identification accuracy as a function of category membership and representativeness of the everyday environment, a repeated-measures analysis of variance (ANOVA) was conducted on proportion correct for each of the six categories. Identification accuracy in terms of proportion correct varied significantly as a function of the six categories, $F(5,70) = 31.43$, $p < 0.001$, ${}_p\eta^2 = .69$. Post-hoc contrasts revealed that alarms and human-generated sounds did not differ significantly from one another, $p = .47$, nor did the vehicle and mechanical categories, $p = .07$. All other categories significantly differed from one another, $p < .05$. The lack of difference between alarms and humans could reflect a benefit of signal salience or prior experience with signals of that type. In the case of both alarms and human-generated sounds, there are both behavioral and neural data to suggest preferential processing (related discussion in Starvropoulos and Carver [2016]). The mechanical and vehicle groups were not significantly different from one another, which could indicate that for the participants tested these signals did not represent distinct categories but rather were examples of members from a single broader category. Practically speaking, the differences or lack thereof among each of the categories could be due to uneven numbers of items at each category or the experimenter-defined categorical structure. (This issue is reviewed in detail in Section 5.)

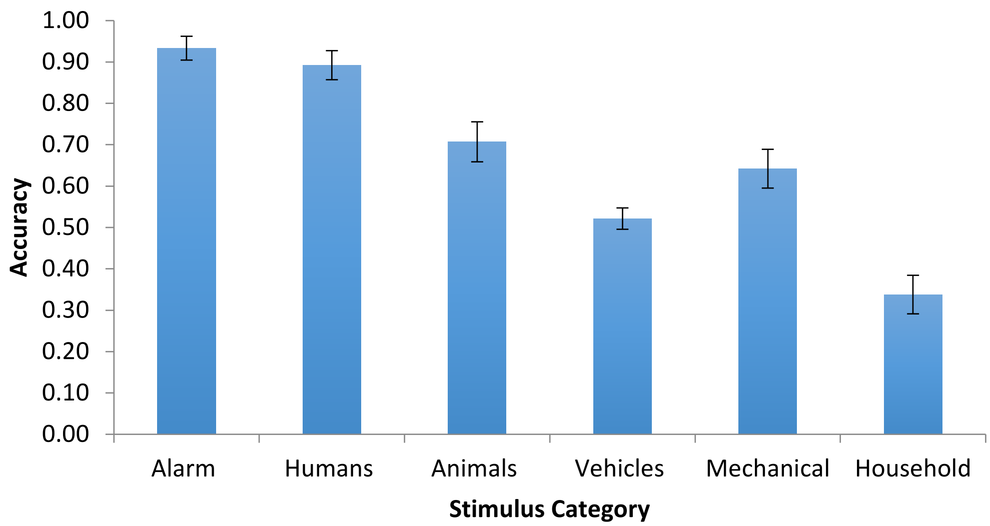

Fig. 1 Data show significant differences in accuracy across different stimulus categories, although superordinate category membership did not explain a significant portion of the variance in response.

4.2 Confidence Ratings

On each trial following the identification response, listeners were asked to indicate on a scale of 1 to 7, from least to most confident, how certain they were of the accuracy of the label they generated for each sound. Figure 2 shows the average confidence rating for each of the sound categories. Overall listener confidence ratings were slightly skewed toward more positive (M = 5.59, SE = 0.13, Min = 3.46, Max = 6.88), suggesting that listeners reported they thought their responses were generally accurate. A repeated-measures ANOVA with category (6) as a factor was conducted for the confidence ratings. The results showed a significant difference in confidence ratings across categories, $F(5, 70) = 21.09$, $p < 0.001$, ${}_{p}\eta^2 = 0.60$. However, despite an overall effect of category, post-hoc contrasts revealed that this difference was primarily driven by differences between two categories. The vehicle and household items categories were significantly different from one another ($p < 0.05$), with lower confidence ratings for the household sound group, and these two categories had significantly lower confidence ratings than all other categories ($p < 0.01$). One qualitative observation from the identification responses associated with the confidence ratings for the items in these two categories in particular: The identification descriptions provided when confidence was low tended to be less specific and rely on creating words to articulate the sound produced by the object, rather than an object label, as was the case for the other categories. For example, a participant who did not use the label “motorcycle” instead wrote “motor going blubbbbbbb”. This anecdote suggests there may be differences in the way the identity of a sound is represented when a label is available versus unavailable. We are currently following up on this effect with a lexical and content analysis for the sounds used here along with a related image set.

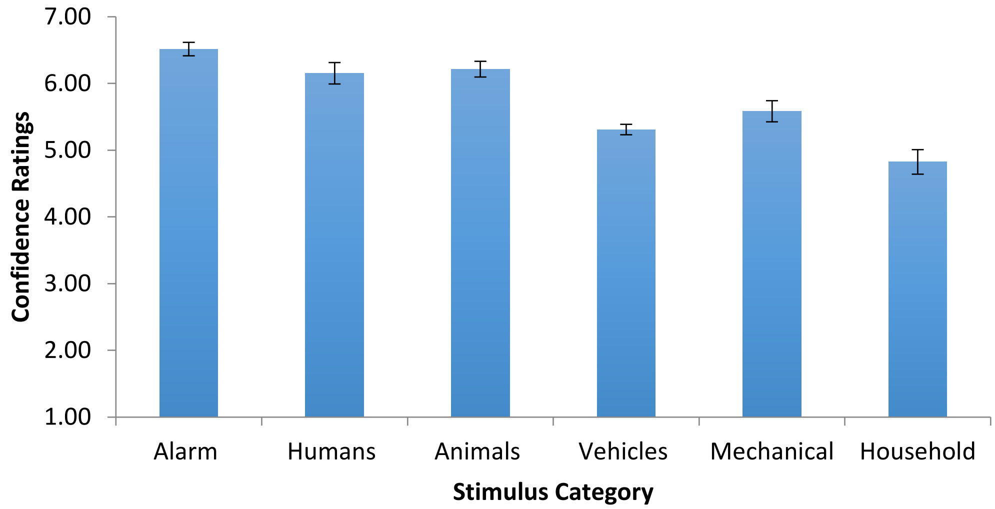

Fig. 2 The effect of stimulus category on response confidence was significant; what is clear from the means is the differences across categories are smaller than those observed for the accuracy measure.

4.3 Relationship between Accuracy and Response Confidence

Overall, there was a positive correlation between identification accuracy (calculated as proportion correct) and confidence in responding $r = 0.56$, $p < 0.001$ (Fig. 3). Multiple regression was used to further explore the relationship among identification accuracy, confidence ratings, and stimulus category. With identification accuracy as the dependent variable, the model was significant $R^2 = 0.353$, $F(2, 612) = 168.70$, $p < 0.001$. These results suggest that both stimulus category and a listener's confidence in their response are important determinants of response accuracy for identification tasks; however, further study is needed to determine a mechanism underlying the link between accuracy and confidence.

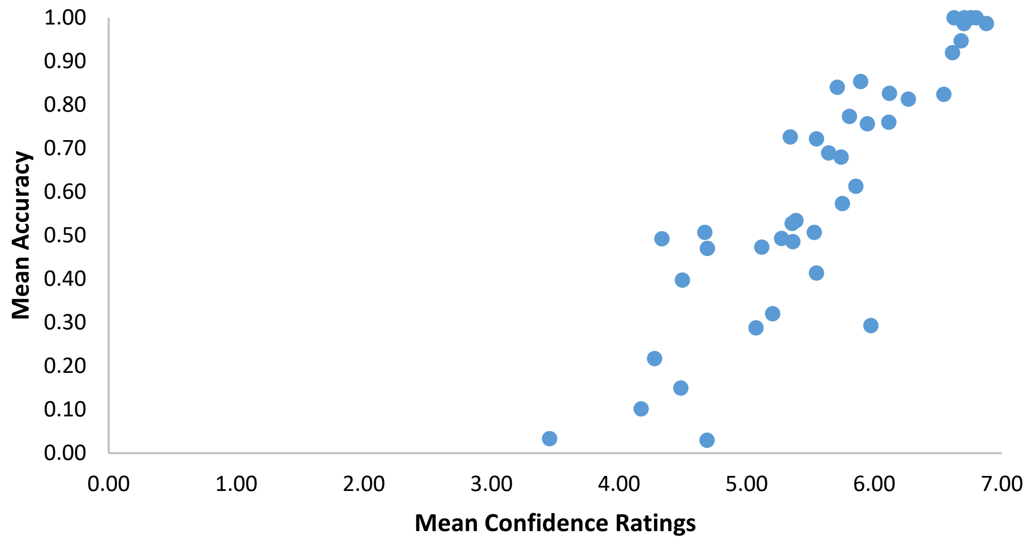

Fig. 3 Strong positive correlation shown between identification accuracy and listeners' confidence in their responses

5. Conclusions

To summarize, the results of the current study reveal that sound-identification accuracy changes as a function of sound category, but that experimenter-defined categories may not best represent the way participants would group sounds together when asked. The benefit of a priori versus post-hoc category formation is a methodological question, and there is no clear "correct" strategy for grouping stimuli. Conservatively, if the goal is stimulus characterization through documentation of the semantic properties of a stimulus set, it may be better to examine the stimuli as individual items rather than categorically. Further, category membership often depends on the context provided by the full set of items, so categories formed during open-ended identification may not generalize to a different response metric or a set of items that contains only a subset of the original items. Thus, while categorization helps streamline the discussion of the findings for the current stimulus set, for stimulus norming purposes it may be "best practice" to consider individual items. Practically, identification of environmental sounds is an important aspect of *environmental awareness* or, in the military domain, *situation awareness* (SA). Environmental sound identification is linked to both the acoustics and the semantics of the sound-producing event. The evidence provide by our open-ended identification task supports this finding; semantic information is critical in supporting good SA. Semantic information supports environmental sound because environmental sounds are inherently meaningful and the differences in identification as a function of concreteness suggest there may be lexical- or linguistic-level processing of these sound events. A number of studies have revealed there are common perceptual processes underlying speech and nonspeech perception (Deihl et al. [2004] has an excellent review) and, specifically, others have suggested that both acoustic and semantic properties contribute to environmental sound perception (Gygi et al. 2007; Giordano et al. 2010). The present study is one in a series of experiments that seeks to better understand and quantify meaning-based or semantic influence on perception for items that are common in real-world environments.

6. References

Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. J Exp Psych Hum Perc Perf. 1993;19(2):250–267.

Ballas JA, Howard JH. Interpreting the language of environmental sounds. Env Beh. 1987;19(1):91–114.

Bergman P, Sköld A, Västfjäll D, Fransson N. Perceptual and emotional categorization of sound. J Acou Soc America. 2009;126(6):3156–3167.

Catchpole KR, McKeown JD, Withington DJ. Localisable auditory warnings: integral 'where' and 'what' components. In: Harris D, editor. Engineering psychology and cognitive ergonomics: vol. 4: job design, product design and human-computer interaction. New York (NY): Routledge. 2017:439–446.

Cohen J. A coefficient of agreement for nominal scales. Ed Psych Meas. 1960;20(1):37–46.

Cummings A, Čeponienė R, Koyama A, Saygin AP, Townsend J, Dick F. Auditory semantic networks for words and natural sounds. Brain Res. 2006;1115(1):92–107.

Diehl RL, Lotto AJ, Holt LL. Speech perception. Ann Rev Psych. 2004;55:149–179.

Duff M, Moser A, Olsen S, Foots A, Gerhardstein P, Dickerson K. Didn't I just see that sound? Cross-modal facilitation during an identification task. Cog Res Princ Imp. Forthcoming 2018.

Foots A, Dickerson K, Gaston J. Characterizing real-world auditory scenes using 360° audio-visual capture. Poster presented at 172nd Meeting of the Acoustical Society of America; 2016 Nov 28–Dec 2; Honolulu, HI.

Giordano BL, McDonnell J, McAdams S. Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds. Brain Cog. 2010;73(1):7–19.

Gregg MK, Samuel AG. The importance of semantics in auditory representations. Att Perc Psychop. 2009;71(3):607–619.

Gygi B, Kidd GR, Watson CS. Similarity and categorization of environmental sounds. Perc Psychop. 2007;69(6):839–855.

Gygi B, Shafiro V. From signal to substance and back: Insights from environmental sound research to auditory display design. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K, editors. Auditory display. Berlin–Heidelberg (Germany): Springer; 2009. p. 306–329.

Gygi B, Shafiro V. Development of the database for environmental sound research and application (DESRA): design, functionality, and retrieval considerations. EURASIP J Aud Speech Music Proc. 2010.

Houix O, Lemaitre G, Misdariis N, Susini P, Urdapilleta I. A lexical analysis of environmental sound categories. J Exp Psych App. 2012;18(1):52–80.

Leech R, Gygi B, Aydelott J, Dick F. Informational factors in identifying environmental sounds in natural auditory scenes. J Acou Soc America. 2009;126(6):3147–3155.

Lemaitre G, Heller LM. Evidence for a basic level in a taxonomy of everyday action sounds. Exp Brain Res. 2013;226(2):253–264.

Lemaitre G, Houix O, Misdariis N, Susini P. Listener expertise and sound identification influence the categorization of environmental sounds. J Exp Psych App. 2013;16(1):16–32.

Marcell MM, Borella D, Greene M, Kerr E, Rogers S. Confrontation naming of environmental sounds. J Clin Exp Neuro. 2000;22(6):830–864.

McArdle J, Foots A, Stachowiak C, Dickerson K. Strategies for characterizing the sensory environment: objective and subjective evaluation methods using the VisiSonic real space 64/5 audio-visual panoramic camera. Aberdeen Proving Ground (MD): Army Research Laboratory (US): 2017 Nov. Report No.: ARL-TR-8205.

Psychology Software Tools, Inc. E-Prime 2.0. 2012. [accessed 2016 Oct] https://www.pstnet.com.

Stavropoulos KKM, Carver LJ. Neural correlates of attention to human-made sounds: an ERP study. PloS ONE. 2016;11(10):e0165745.

Truax B. Acoustic communication. Vol. 1. Westport (CT): Greenwood Publishing Group; 2001.

VanDerveer NJ. Confusion errors in identification of environmental sounds. J Acou Soc America. 1979;65(S1):S60–S60.

Van Petten C, Rheinfelder H. Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. Neuropsychologia. 1995;33(4):485–508.

List of Symbols, Abbreviations, and Acronyms

ANOVA	analysis of variance
ERP	event-related potential
SA	situation awareness
SE	standard error
XAFC	multiple-alternative forced choice

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

2 (PDF) DIR ARL
IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 (PDF) GOVT PRINTG OFC
A MALHOTRA

1 (PDF) ARL
RDRL HRB B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

8 (PDF) ARL
SFC PAUL RAY SMITH CENTER
RDRL HRO COL H BUHL
RDRL HRF J CHEN
RDRL HRA I MARTINEZ
RDRL HRR R SOTTILARE
RDRL HRA C A RODRIGUEZ
RDRL HRA B G GOODWIN
RDRL HRA A C METEVIER
RDRL HRA D B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 (PDF) USA ARMY G1
DAPE HSI B KNAPP
300 ARMY PENTAGON
RM 2C489
WASHINGTON DC 20310-0300

1 (PDF) USAF 711 HPW
711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 (PDF) USN ONR
ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 (PDF) USA NSRDEC
RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 (PDF) OSD OUSD ATL
HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

12 (PDF) ARL
RDRL HR
J LOCKETT
P FRANASZCZUK
K MCDOWELL
K OIE
RDRL HRB
D HEADLEY
RDRL HRB C
J GRYNOVICKI
RDRL HRB D
C PAULILLO
RDRL HRF A
A DECOSTANZA
RDRL HRF B
A EVANS
RDRL HRF C
J GASTON
RDRL HRF D
A MARATHE
K DICKERSON